AD_____

Award Number: DAMD17-02-1-0246

TITLE: A Large Scale Genomic Approach to Prostate Cancer Risk in African American Men

PRINCIPAL INVESTIGATOR: Matthew L. Freedman, M.D.

CONTRACTING ORGANIZATION: Massachusetts General Hospital
Boston, Massachusetts 02114

REPORT DATE: June 2003

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 2003	Annual Summary (1 Jun 2002 - 31 May 2003)

4. TITLE AND SUBTITLE
A Large Scale Genomic Approach to Prostate Cancer Risk in African American Men

5. FUNDING NUMBERS
DAMD17-02-1-0246

6. AUTHOR(S)
Matthew L. Freedman, M.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Massachusetts General Hospital
Boston, Massachusetts 02114

E-Mail: mfreedman@partners.org

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Original contains color plates: All DTIC reproductions will be in black and white.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

We are using a haplotype based approach to understand how genetic variation in the GH/IGF-1 pathway affects prostate cancer risk in African-American men. The sequencing of the human genome along with the characterization and cataloguing of single nucleotide polymorphisms (SNPs) provides the raw material to understand how genotypic variation translates into phenotypic variation including disease risk. While prostate cancer is known to have a substantial genetic contribution to risk, progress has been slow in identifying the genes involved in the disease. By testing a large population in an association study design, we are powered to detect the modest risks that may be involved in this disease.

14. SUBJECT TERMS
Genomics, Genetics, Prostate Cancer Risk

15. NUMBER OF PAGES
12

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION:

Prostate cancer is the most common non-cutaneous cancer in North American men. Moreover, African-American men suffer from the highest measured incidence of prostate cancer in the world. This increased risk is, in part, due to genetic factors [1]. IGF-1 serum levels have, in fact, been implicated in prostate cancer pathogenesis. We chose to focus on understanding the genotype phenotype relationships in 39 genes involved in the IGF-1 pathway and prostate cancer risk in a large multiethnic cohort. Our plan is to systematically examine common (>.05) genetic variation in the IGF-1 pathway. By focusing on coding as well as noncoding variation, we can insure that we will comprehensively capture the bulk of variation in the loci under consideration. Assessment of the coding regions will be performed by deep resequencing in 96 individuals, while the noncoding regions will be subjected to a haplotype based analysis. Once this genetic variation is characterized and catalogued, we propose to genotype the polymorphisms in a multiethnic case-cohort study consisting of N=4,497 , spanning four self-reported ethnic groups (African-American, Caucasian, Japanese and Latino). We can then look for associations between certain variants and the risk of developing prostate cancer.

We also proposed to measure and control for population stratification, if present. Population stratification is a source of confounding that can potentially lead to false positive results due to allele frequency differences between cases and controls at loci throughout the genome. While many theoretical debates have surfaced over the years, only recently has it become possible to empirically study this potential source of confounding.

BODY:

Task 1. Evaluation of possible cryptic ethnic stratification in case and control populations in order to eliminate this potential source of false positives.

We have exceeded the number of markers and evaluation of stratification as outlined in Task 1 in the Statement of Work in order to thoroughly study this topic. We intensively studied the African American population based on the hypothesis that stratification would be more likely to occur in this group than other populations since the prevalence of prostate cancer is higher in Africans than Europeans [2]. Hence, African-American cases would be expected to possess a higher proportion of African ancestry than controls, leading to systematic differences in allele frequencies. Initially, we genotyped 46 markers in 93 cases and 86 controls to assess for stratification (Table 1). A summary χ^2 statistic revealed no significant stratification [3]. However, using a more quantitative metric, termed genomic control (GC), the data were still consistent with stratification [4]. This could be due to one of two scenarios: a) cryptic stratification is present (subtle degrees of stratification that are not adequately captured by self-reported ethnicity), or b) stratification is not present.

To differentiate between these alternatives, we increased our power to detect stratification by genotyping 138 markers in our sample of 467 African-American prostate cancer cases and 512 controls. We discovered that statistically significant stratification was, in fact, present in this study. Notably, this effect is present in a case-cohort designed study, which should be less susceptible to the effects of stratification. Although the magnitude of this effect may seem modest, it is expected to impact the false positive rate of a study, especially when trying to identify genetic variants that confer risk in a complex disease such as cancer. For example, with the estimated upper bound on λ of 2.25 in a study of 500 cases / controls, we would expect that if 100 hypotheses were tested, two false-positive results would be observed (a 2% false-positive rate).

Our analysis reveals that population stratification affects case-control studies in practice, and that despite the uttermost care in matching cases and controls, it is likely to become increasingly important factor in case-control studies of the future, as sample sizes increase in order to detect more subtle genetic effects and correct for multiple hypothesis testing. Importantly, Genomic Control provides a safe way to preserve the power of case-control studies while controlling for this source of false-positives.

Tasks 2 and 3: Obtain genetic variation information, in the form of SNPs, for each of the 39 genes to be investigated in the growth hormone pathway and evaluate marker assays. Assess genetic variation in genes in growth hormonal pathway.

We are continuing to make progress on this front. We have learned a great deal through our study of the androgen receptor (AR). We have completed data collection and analysis and are in the process of preparing the manuscript for submission. Our strategy included sequencing the coding region for missense variants, typing the well-described

CAG microsatellite in exon 1 as well as a haplotype analysis to cover the noncoding portion of the locus (see Figure 1).

We resequenced the exons in 88 advanced cases of prostate cancer and did not find any amino acid altering variants. We genotyped the CAG microsatellite polymorphism in 2,266 cases and controls and found a nominally significant association when analyzing this repeat as a continuous variable consistent with prior reports in the literature [5, 6]. To survey the noncoding region, we genotyped a total of 32 polymorphic SNPs across ~275 kb in a multiethnic population (African-American, Caucasian, Japanese and Latino). The AR can be described by three blocks of extensive linkage disequilibrium. We used the strict criteria as set forth in Gabriel et al., to define a block [7]. As seen in prior studies, the African-American population possesses the greatest diversity, i.e., 30 polymorphic markers. In sharp contrast, the Japanese population is monomorphic at all 32 sites. Thus, while the AA population has 14 haplotypes across this region, only 1 haplotype is segregating in the Japanese population. We tested these haplotypes in a large prostate cancer cohort (African-American, N=1,003, Caucasian, N=209, Japanese, N=242, and Latino, N=302). The haplotypes did not reveal any evidence of association with prostate cancer risk.

We have adopted the following process to characterize variants that we find in the databases. We work with a "haplotype" plate that reflects the ethnic composition in the prostate cohort. This plate that contains approximately 70 individuals of each ancestral origin outlined above. By testing our markers in a smaller independent population, we are able to discern which markers are monomorphic in a given population as well as which assays do not work. Most importantly, we are able to define blocks and haplotypes in these plates. As there are usually many more markers that are typed than are necessary to describe haplotype variation, we are able to use a computer program designed by our collaborators (D. Stram, USC, manuscript submitted), to efficiently use the minimal number of SNPs that captures the full spectrum of variation. This process allows a streamlined, efficient method of characterizing SNPs and haplotypes.

For the first year, we have prioritized the top candidates in the IGF-1 pathway based on epidemiologic data [8-11]. We have currently completed or in the process of completing the haplotype structure in 5 genes, PIK3CG, IGF1, IGFBP1, IGFBP3, and IGF2R (Figures 2 and 3). We have tremendously optimized workflow over the past year and the SNP databases are much more complete allowing a more thorough characterization of the loci.

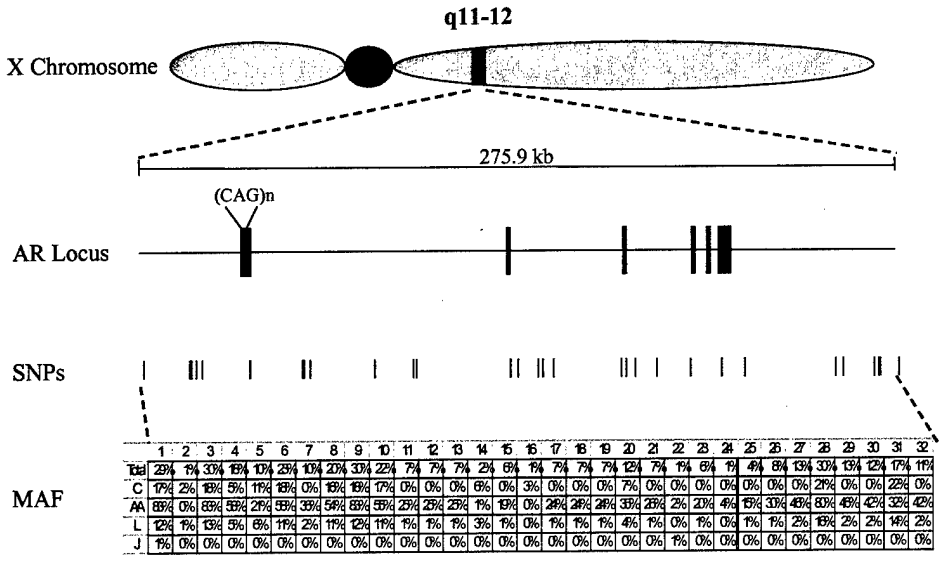

Fig. 1

Figure 1: Overview of androgen receptor locus showing SNPs typed and minor allele frequencies in the different populations.

This figure demonstrates a schematic outline showing the relationship of SNPs to the AR locus. The allele frequencies are shown for each ancestral population. Notably, all of the SNPs typed in the Japanese population are monomorphic, whereas the majority of SNPs are polymorphic in the African American population.

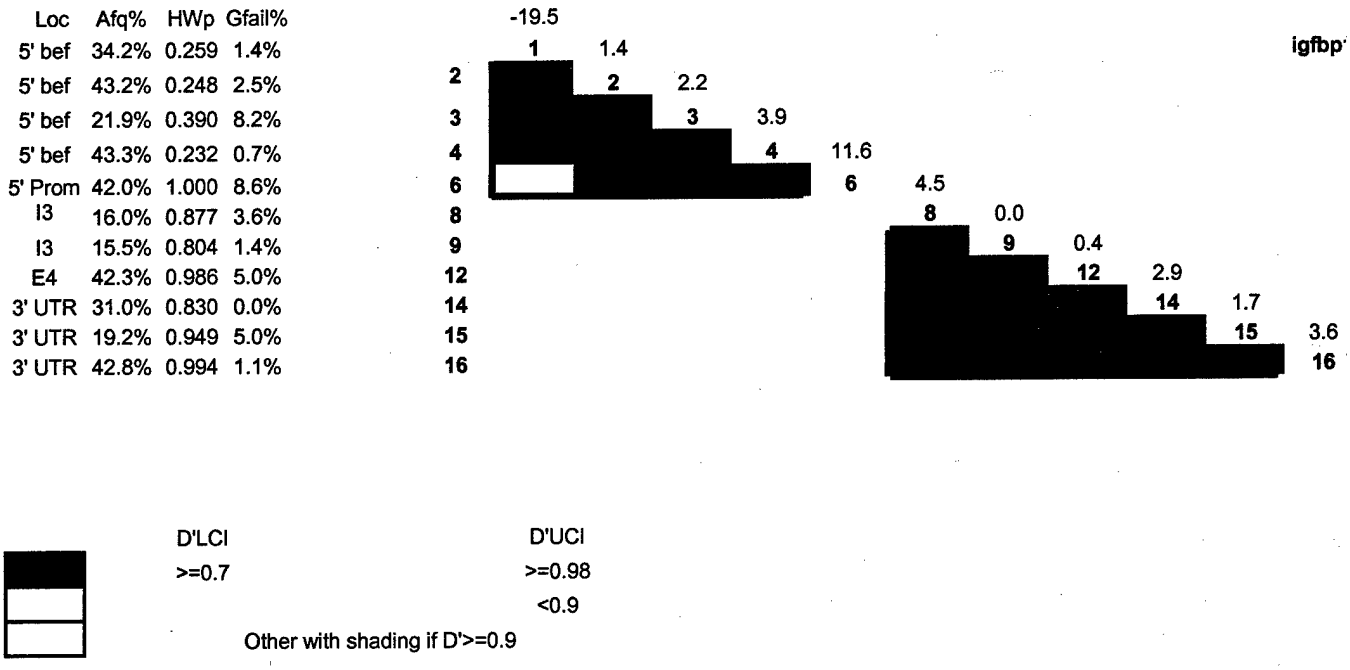

Figure 2: D prime plot of IGFBP-1
This figure demonstrates the linkage disequilibrium relationships for all SNPs greater than 10% in order to define a block. D' is a pariwise measure of linkage disequilibrium. Blocks were defined according to Gabriel et al ***CITE***. The colors in the plot correspond to the confidence limits on the estimate of D' (see Gabriel et al.). A red square denotes a high D' estimate with tight confidence intervals, the white square is a lower confidence interval. The chart on the left corresponds to the location of the SNP, the allele freq (Afq%), the Hardy-Weinberg p value and the % failure (Gfail%). Two blocks are present in this gene that are highlighted by the black lines (SNPs 1-6 and SNPs 8-16)

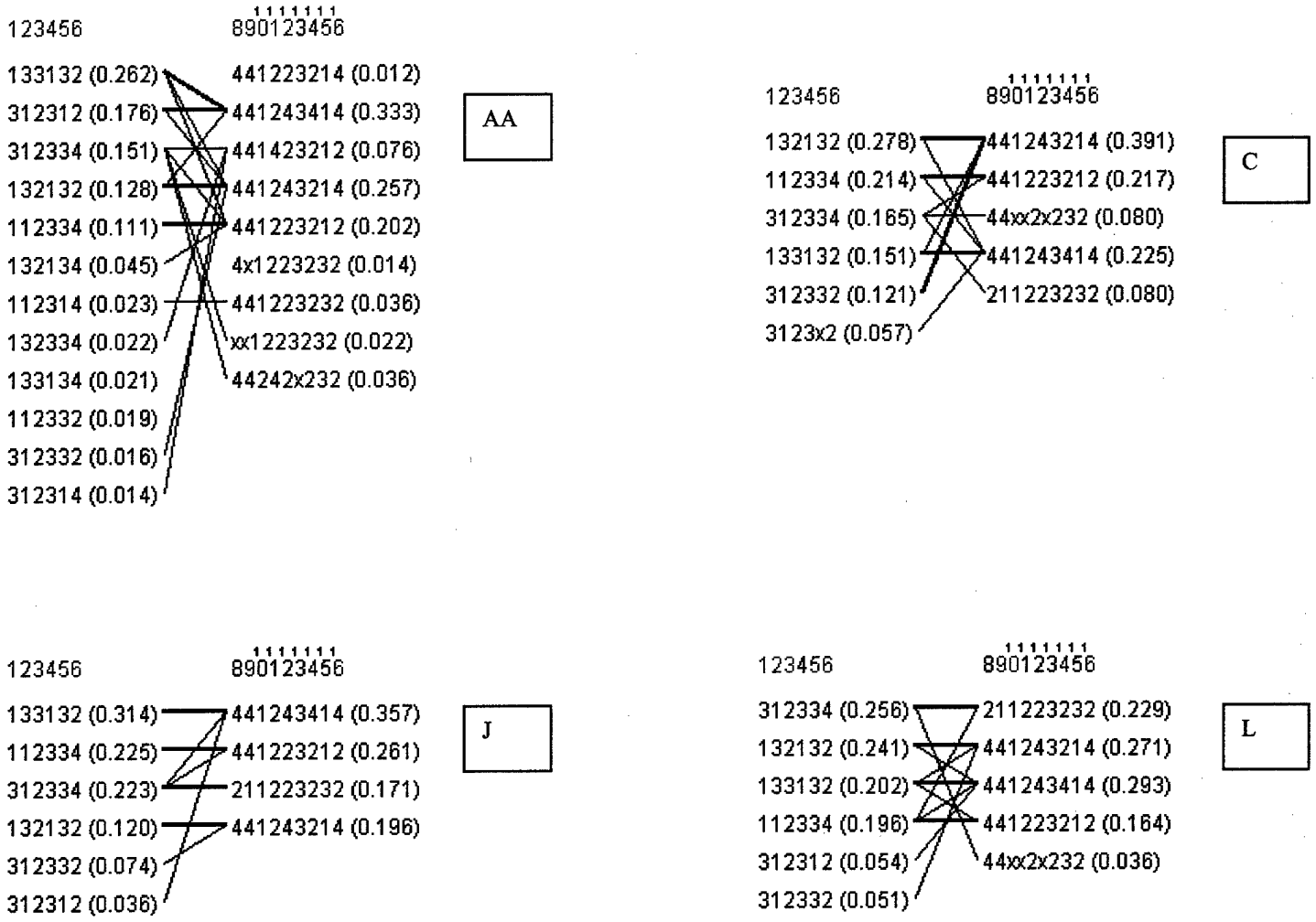

Figure 3: Haplotype structure of IGFBP-1 in a Multiethnic cohort.

This diagram depicts the haplotypes within each block for for IGFBP-1. The numbers on top (in red and black) refer to the SNP number (they correspond to the same SNPs as in Figure 1). Each haplotype is coded in numbers (1=A, 2=C, 3=G, 4=T). The haplotypes are estimated using a standard EM algorithm and the associated frequencies are shown in parentheses. The haplotypes are shown by self reported ethnicity (AA=African-American, C=Caucasian, J=Japanese, L=Latino). The lines between the blocks demonstrate recombination events. As can be seen, the vast majority of haplotypes are shared between each population.

KEY RESEARCH ACCOMPLISHMENTS:

1. Hiring of stellar technician.

2. Providing basis for understanding the impact of stratification on case control studies likely to be used in future association studies.

3. Androgen Receptor

    a. Further understanding of range of human diversity in multi-ethnic populations at this locus.
    b. Nominally significant association between CAG microsatellite repeat in exon 1 and prostate cancer risk.
    c. No association between AR haplotypes and prostate cancer risk.

4. Haplotype structure determined for IGF1, IGFBP1, IGFBP3, PIK3CG, and IGF2R in a multiethnic population.

REPORTABLE OUTCOMES:

1. Abstract Presentation on IGF-1 at 2003 American Association for Cancer Research Annual Meeting

2. Manuscript in preparation for stratification data.

3. Manuscript in preparation for androgen receptor data.

CONCLUSIONS:

1. We provide evidence that even subtle amounts of stratification can lead to false positive outcomes in a large association study looking for modest genetic effects. This conclusion stands in contrast to beliefs that stratification bias will not affect an association study if cases and controls are carefully matched using self-reported ethnicity as a proxy for ancestry. We provide an approach to test and conservatively correct for this source of confounding.

2. The androgen receptor data reveals striking differences in haplotype frequencies between different populations. These observations are extremely important, especially when studying diseases such as prostate cancer that demonstrate clear ethnic predispositions.

3. Will test haplotypes in large prostate cancer cohort to determine if any variants confer risk of prostate cancer in large population.

REFERENCES:

1. H, G., *Prostate cancer epidemiology.* Lancet, 2003. **361**(9360): p. 859-64.
2. Bunker CH, P.A., Konety BR, Dhir R, Brufsky AM, Vivas CA, Becich MJ, Trump DL, Kuller LH, *High prevalence of screening-detected prostate cancer among Afro-Caribbeans: the Tobago Prostate Cancer Survey.* Cancer Epidemiol Biomarkers Prev, 2002. **11**(8): p. 726-9.
3. Pritchard JK, R.N., *Use of unlinked genetic markers to detect population stratification in association studies.* Am J Hum Genet, 1999. **65**(1): p. 220-8.
4. Reich DE, G.D., *Detecting association in a case-control study while correcting for population stratification.* Genet Epidemiol, 2001. **20**(1): p. 4-16.
5. Hsing, A.W., et al., *Polymorphic CAG and GGN repeat lengths in the androgen receptor gene and prostate cancer risk: a population-based case-control study in China.* Cancer Res, 2000. **60**(18): p. 5111-6.
6. Giovannucci, E., et al., *The CAG repeat within the androgen receptor gene and its relationship to prostate cancer.* Proc Natl Acad Sci U S A, 1997. **94**(7): p. 3320-3.
7. Gabriel SB, S.S., Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D, *The structure of haplotype blocks in the human genome.* Science, 2002. **296**(5576): p. 2225-9.
8. Chokkalingam AP, P.M., Fillmore CM, Gao YT, Stanczyk FZ, Deng J, Sesterhenn IA, Mostofi FK, Fears TR, Madigan MP, Ziegler RG, Fraumeni JF Jr, Hsing AW, *Insulin-like growth factors and prostate cancer: a population-based case-control study in China.* Cancer Epidemiol Biomarkers Prev, 2001. **10**(5): p. 421-7.
9. Chan JM, S.M., Ma J, Gann P, Gaziano JM, Pollak M, Giovannucci E, *Insulin-like growth factor-I (IGF-I) and IGF binding protein-3 as predictors of advanced-stage prostate cancer.* J Natl Cancer Inst, 2002. **94**(14): p. 1099-106.
10. Chan JM, S.M., Giovannucci E, Gann PH, Ma J, Wilkinson P, Hennekens CH, Pollak M, *Plasma insulin-like growth factor-I and prostate cancer risk: a prospective study.* Science, 1998. **279**(5350): p. 563-6.
11. Stattin P, B.A., Rinaldi S, Biessy C, Dechaud H, Stenman UH, Egevad L, Riboli E, Hallmans G, Kaaks R, *Plasma insulin-like growth factor-I, insulin-like growth factor-binding proteins, and prostate cancer risk: a prospective study.* J Natl Cancer Inst, 2000. **92**(23): p. 1910-7.